Article

The genomic natural history of the aurochs

https://doi.org/10.1038/s41586-024-08112-6

Received: 3 May 2024

Accepted: 25 September 2024

Published online: 30 October 2024

Check for updates

Conor Rossi[1], Mikkel-Holger S. Sinding[2]✉, Victoria E. Mullin[1], Amelie Scheu[1,3], Jolijn A. M. Erven[1,4], Marta Pereira Verdugo[1], Kevin G. Daly[1,5], Marta Maria Ciucani[6], Valeria Mattiangeli[1], Matthew D. Teasdale[1,7], Deborah Diquelou[1], Aurélie Manin[8], Pernille Bangsgaard[6], Matthew Collins[6,9], Tom C. Lord[10], Viktor Zeibert[11,41], Roberto Zorzin[12], Michael Vinter[13], Zena Timmons[14], Andrew C. Kitchener[14,15], Martin Street[16], Ashleigh F. Haruda[8], Kristina Tabbada[8], Greger Larson[8], Laurent A. F. Frantz[17,18], Birgit Gehlen[19], Francesca Alhaique[20], Antonio Tagliacozzo[20], Mariagabriella Fornasiero[21], Luca Pandolfi[22], Nadezhda Karastoyanova[23], Lasse Sørensen[24], Kirill Kiryushin[25], Jonas Ekström[26], Maria Mostadius[26], Aurora Grandal-d'Anglade[27], Amalia Vidal-Gorosquieta[27], Norbert Benecke[28], Claus Kropp[29], Sergei P. Grushin[30], M. Thomas P. Gilbert[6], Ilja Merts[31], Viktor Merts[31], Alan K. Outram[32], Erika Rosengren[33,34,35], Pavel Kosintsev[36,37], Mikhail Sablin[38], Alexey A. Tishkin[30], Cheryl A. Makarewicz[39,40], Joachim Burger[3] & Daniel G. Bradley[1]✉

Now extinct, the aurochs (*Bos primigenius*) was a keystone species in prehistoric Eurasian and North African ecosystems, and the progenitor of cattle (*Bos taurus*), domesticates that have provided people with food and labour for millennia[1]. Here we analysed 38 ancient genomes and found 4 distinct population ancestries in the aurochs—European, Southwest Asian, North Asian and South Asian—each of which has dynamic trajectories that have responded to changes in climate and human influence. Similarly to *Homo heidelbergensis*, aurochsen first entered Europe around 650 thousand years ago[2], but early populations left only trace ancestry, with both North Asian and European *B. primigenius* genomes coalescing during the most recent glaciation. North Asian and European populations then appear separated until mixing after the climate amelioration of the early Holocene. European aurochsen endured the more severe bottleneck during the Last Glacial Maximum, retreating to southern refugia before recolonizing from Iberia. Domestication involved the capture of a small number of individuals from the Southwest Asian aurochs population, followed by early and pervasive male-mediated admixture involving each ancestral strain of aurochs after domestic stocks dispersed beyond their cradle of origin.

Now extinct, the aurochs (*Bos primigenius*) ranged widely in Eurasia and North Africa. Its fossil presence in Europe stretched from 650 thousand years ago (ka)[2], but its earliest probable ancestors are detected in South Asia[3], close to the centre of diversity for the tribe Bovini, to which all wild cattle belong. During the Holocene, the aurochs was the largest mammal present in temperate fertile zones, and its grazing was instrumental in the persistence of mixed forest and grassland[4]. Because humans have thrived in similar ecosystems, the aurochs has been intertwined with human culture and society since the Upper Pleistocene; first as a hunted prey, but also as a rich source of iconography, stretching from Magdalenian cave art through to the carvings at Göbekli Tepe, and the bucrania of Çatalhüyük[1]. This relationship persisted until its demise with the death of the last cow in AD 1627 in Poland. The last bull had been shot seven years earlier and its horns, like those of many others, were fashioned into royal hunting horns or ceremonial drinking vessels[5]. Today, the aurochs' descendants, domestic cattle (*Bos taurus*), comprise approximately one-third of Earth's mammalian biomass[6].

Bos primigenius was one of the first successful targets for ancient DNA research; early work showed that the mitochondrial DNA (mtDNA) of European aurochs and domestic cattle was divergent[7,8]. More recently, analyses of four autosomal genomes have supported the inference of interbreeding between aurochsen and cattle in Europe, the Levant and Africa[9–11]. Here we report an analysis of genomes from 38 (33 newly reported) aurochs specimens sampled across Eurasia (spanning from more than 47,000 calibrated years before present (cal. BP) to 4,000 cal. BP; 0.06–22.47×, mean 3.9× genomic depth), in addition to new sequence data from 8 Central Asian ancient cattle genomes (spanning from 6,400 cal. BP to 3,800 cal. BP; 0.44–2.43×, mean 1.17×).

Dynamic genetic geography of the aurochs

A plot summarizing genomic affinities (Fig. 1c) highlights three poles of *B. primigenius* diversity. On the top left are those genomes from Europe (green); on the right three divergent North Asian genomes dating from the Upper Pleistocene (pink); and to the bottom left two Southwest Asian aurochsen (brown), which cluster with ancient domestic cattle and a North African aurochs. *Bos indicus* are more divergent and are not plotted. The trajectories of the population ancestries embedded in these clusters were dynamic and responded to both climatic transition and human management after domestication.

In Europe, the clustering of two pre-Last Glacial Maximum (LGM; 27–18 ka) European aurochsen (46 ka, 47+ ka) together with later

A list of affiliations appears at the end of the paper.

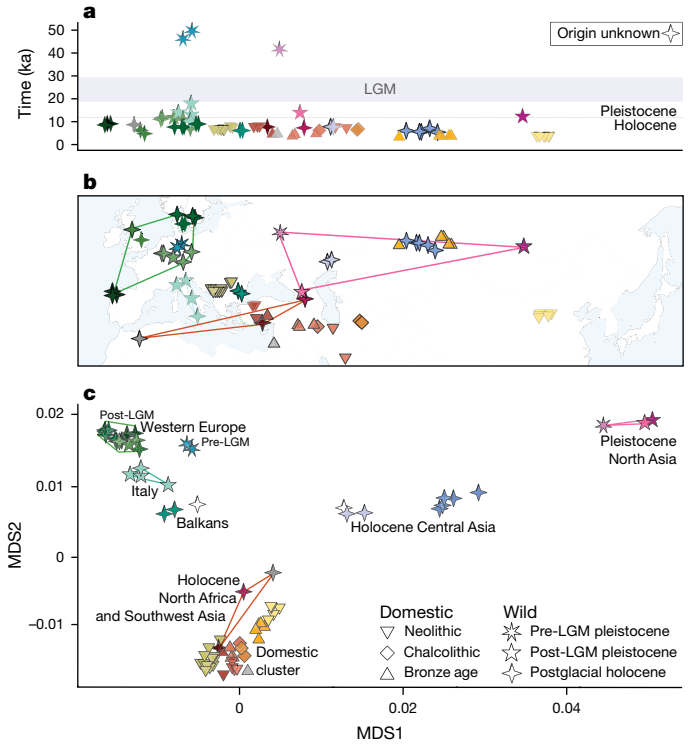


Fig. 1 | Temporal, geographical and genetic location of 72 ancient *Bos* genomes. a,**b**, Temporal (**a**) and spatial (**b**) provenance of 38 aurochsen (*Bos primigenius*) and 34 ancient cattle (*Bos taurus*) genomes included in this study (*Bos indicus* specimens are not plotted). **c**, Genetic locations, using multidimensional scaling (MDS) of pairwise distances based on allele sharing. Symbol colour and shape denote provenance, age and wild or domestic identity of individuals (aurochsen from different time periods are indicated by three different star symbols). These are held constant across figure panels, illustrating the correspondence between genetics and geography. The major clusters of aurochsen selected for discussion (Pleistocene North Asia; Holocene North Africa and Southwest Asia; and several European groups) are demarcated by lines on both the map and the MDS plot, and are labelled in **c**.

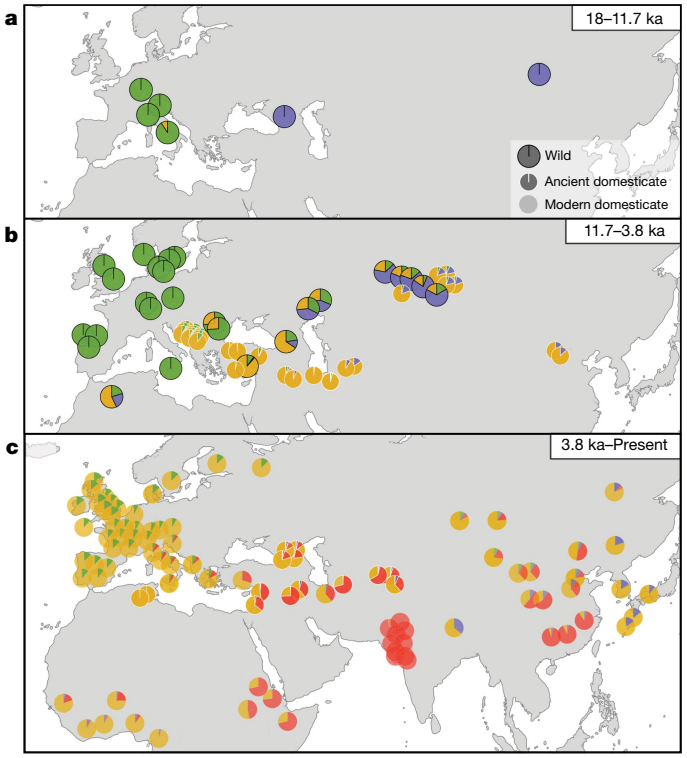


Fig. 2 | Plots of ancestral components in aurochs and domestic cattle genomes. a–**c**, Plots of ancestral population components inferred among post-LGM aurochsen and cattle using NGSadmix. Individual aurochs samples are denoted by larger circles and cattle by smaller circles without borders. Optimizing at $K = 4$, four interpretable components emerge differently distributed in each time interval: European aurochs (green), North Asian aurochs (purple), Southwest Asian aurochs–*Bos taurus* (yellow) and *Bos indicus* (red). **a**, In the first interval, the North Asian and European ancestries appear separate. **b**, In the second interval, during the Holocene these have admixed; also domestic cattle (smaller circles) have emerged and migrated from Southwest Asia, admixing with both European and North Asian aurochs. **c**, The final plot shows the persistence of these latter admixtures as well as the spreading influence of *B. indicus* ancestry in warmer regions.

genomes implies a degree of temporal continuity (Fig. 1c). However, during the LGM, much of central Europe was covered in permafrost tundra[12] and the aurochs' habitat contracted to southern refugia in Franco-Iberia, Italy and the Balkans[13,14]. Genomes from these three regions separate into groups distinguished by different levels of minor admixture with a Southwest Asian aurochs population (Extended Data Fig. 1). Notably, the bulk of other sampled European (central European, Scandinavian and British) genomes cluster tightly with the Iberian group (Fig. 1). This pattern suggests a genetic distinction between refugia, postglacial recolonization of western Europe derived from the Franco-Iberian refugium, and that the Alps are likely to have formed a barrier to northward gene flow from the Italian refugium.

On a broader scale, the most divergent cluster, which we term North Asian, comprises three *B. primigenius* specimens from the Pleistocene (Fig. 1). This group had a wide geographical and temporal provenance that stretched from western Russia (41.5 ka) and Armenia (13.9 ka) to Lake Baikal (12.2 ka) in the east. Later during the Holocene, Central Asian genomes shift towards the centre of the plot (Fig. 1), which suggests that they underwent admixture with more western aurochs populations—an assertion that is confirmed by admixture modelling using qpAdm (Supplementary Information, section 3) and by unsupervised ancestry component estimates in NGSadmix analysis (Fig. 2b).

The third pole includes two Southwest Asian aurochsen from central Türkiye (7.6 ka) and Armenia (7.1 ka). These were proximal in space and time to the domestication processes that were underway in the Fertile Crescent by the mid-eleventh millennium cal. BP. Furthermore, despite its distant provenance, the sole African (Moroccan 8.8 ka) aurochs shares affinity with this cluster. This is confirmed by phylogenetic analysis (Supplementary Information, section 1) and implies ecological continuity along the south Mediterranean littoral. African aurochsen underwent a reduction in size and robusticity in the Holocene, perhaps reflecting replacement with smaller stock from the Levant[15]. Ancient domestic (*B. taurus*) individuals, although sampled widely, also plot near this Southwest Asian cluster (Fig. 1), reflecting their initial recruitment from wild stock in that region around 11 ka and subsequent dispersal with migrating Neolithic farmers[10]. This affinity is also clear from NGSadmix ancestry modelling as ancient *B. taurus* specimens are predominantly modelled with the Southwest Asian aurochs component (coloured yellow; Fig. 2b,c).

European archaic introgression

Aurochsen first appeared in Europe around 650 ka, about the time of the arrival of *Homo heidelbergensis* and Acheulean technologies[16–18]. However, a much more recent coalescence of all our sampled genomes is estimated by a range of analyses (Fig. 3 and Extended Data Figs. 2 and 3). These are: Bayesian analysis of ^{14}C-dated mtDNA (implemented using BEAST2; Fig. 3b) and both genome-wide coalescence with cross-population analysis and pseudodiploid X chromosome analysis

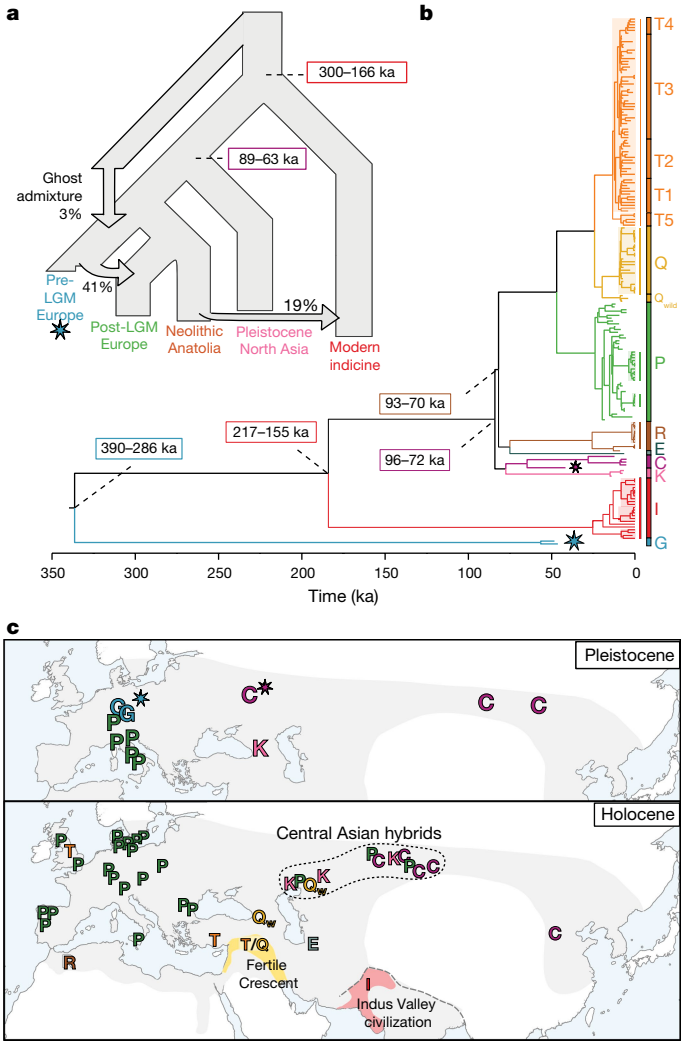


Fig. 3 | Phylogenies of whole genomes and mtDNA. a, AdmixtureGraph-derived phylogeny of the major aurochs genome groups with admixture features common to the highest likelihood solutions. These include a trace (3%) input from an inferred basal population into pre-LGM European aurochsen (marked with asterisks in each panel). Branch divergence times are derived from MSMC2 coalescence analysis (Extended Data Fig. 2). **b**, Bayesian phylogeny of aurochsen and cattle (selected to include rare haplogroups) whole mtDNA sequences; sources of modern sequences are detailed in Supplementary Data 4 and haplogroup nomenclature builds on that of a previous report[25]. Branch time estimates arise from the internal calibration afforded by ^{14}C-dated ancient samples. Haplogroups are labelled on the right and shading and narrow lines denote large clusters with shallow coalescence encountered in domestic cattle; wild mtDNA node coalescences are more distributed in time. **c**, Locations of aurochs haplotypes encountered during the Pleistocene and Holocene; P and G are found in European Pleistocene samples and C and K in North Asian Pleistocene samples. The P haplogroup predominates in European aurochsen sampled during the Holocene, whereas four haplogroups are sampled in Central Asia (P, Q_{wild}, K and C). Q_{wild} refers to haplotypes that have affinity to, but do not fall within, the major Q domestic mtDNA clade. The major domesticated haplogroups (T/Q and I) are inferred to have been present in aurochsen in the Fertile Crescent and Indus Valley.

using the observed bovine mutation rate (implemented using MSMC2; Extended Data Figs. 2 and 3 and Supplementary Information, section 4). These reveal two deep splits within the sampled *Bos* genomes. First, *B. primigenius* versus *B. indicus* (used as a proxy for the South Asian aurochs, *Bos namadicus*) genomes diverged 166–300 ka and mtDNA 217–155 ka; X chromosome coalescence revealed that X chromosomes began to diverge at least 200 ka. Second, within *B. primigenius*, North Asian genomes diverged from European and Southwest Asian genomes 89–63 ka. This is mirrored by their X chromosome divergences, which began to split 85–101 ka. Similarly, North Asian mtDNA variation is characterized by two sister haplogroups, C and K, which diverged from the other major *B. primigenius* haplogroups (P, Q, T, E and R; Fig. 3b,c and Extended Data Fig. 3) 96–72 ka.

It is notable that, despite a much older fossil presence, European, North Asian and Southwest Asian *B. primigenius* (including their descendant *B. taurus*) derived from a common ancestral population within the last 100 ka. However, both genome-wide analyses and mtDNA reveal evidence that earlier archaic European populations made trace contributions to Late Pleistocene genomes. First, whole-genome analysis F_4 tests of tree integrity indicate that a simple bifurcating phylogeny is an incomplete model of aurochs ancestry (Supplementary Information, section 3). Accordingly, Fig. 3a summarizes a best-fitting model estimated using AdmixtureGraph, allowing admixture edges. Our two oldest genomes (>46 ka), from Germany, cluster with other later European genomes, but the model requires an additional minor (3%) input into these pre-LGM genomes from an unsampled basal population. Second, these individuals also have exotic, highly divergent bovine mtDNA chromosomes (haplogroup G; Fig. 3b), which are an outgroup to those of all sampled aurochsen and cattle, including *B. indicus*, joining at a common ancestor timed at up to 390 ka.

Separation during the last glaciation

The estimates for the last common ancestor of North Asian and European–Southwest Asian *B. primigenius* overlapped the latter part of marine isotope stage (MIS) 5, which included milder interstadial stages 5a and 5c (refs. 19–21), during which east–west connectivity might have been facilitated (Extended Data Fig. 3). The subsequent divergence between eastern and western populations coincided with cold and dry stadial periods which persisted from around 71 ka until the Holocene[21]. The three North Asian aurochsen, sampled 41.5–12.2 ka within this cold interval, show unadmixed ancestry profiles (Fig. 2 and Extended Data Fig. 4), perhaps owing to ecological barriers to Europe[22]. During this period, temperate European forests were replaced with open vegetation and steppe[23], which caused range contraction; for example, aurochsen were absent from the Southern Urals, where populations of cold-adapted *Bison priscus* prevailed[24].

The unadmixed ancestry profile of these cold period genomes contrasts with that of later aurochsen from the middle Holocene Caucasus and Central Asia, where admixture is evident in both ancestry profiles and introgression of the western mtDNA haplotypes P and Q (Figs. 2b and 3c and Extended Data Fig. 3). This gene flow was probably enabled by more favourable conditions in the Holocene.

Divergent demography in the aurochs

A decline in effective population size has been observed in several mammalian species, such as Alpine ibex, brown bears and collared lemmings[25–27], possibly associated with the harsher climate during the last glacial period. Using high-coverage (average coverage 8× or higher) samples, and the pairwise serial Markovian coalescent (PSMC) we estimated demographic trajectories through time in 9 ancient and 20 modern *Bos* genomes (Supplementary Information, section 4; selected genomes are plotted in Fig. 4a). We observe that the *B. indicus* ancestor underwent an emphatic inferred population increase. These herds inhabited the probable southern Asian centre of origin for both aurochs and other related Bovini species. This peak might reflect a more benign habitat but could also have been the result of diversity introduced by ancient admixture with an unidentified sister species[28]. Genome analysis of wild cattle relatives from southern and Southeast Asia suggests that secondary introgressions were a feature of Bovini

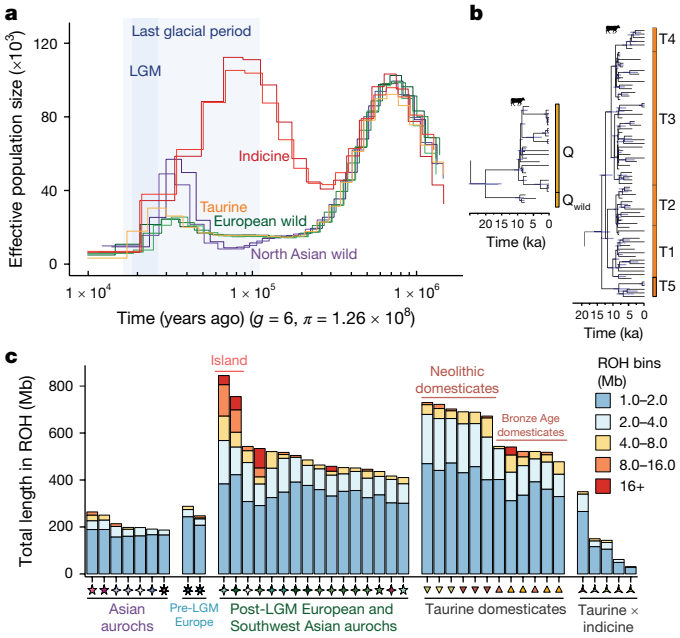


Fig. 4 | Estimation of the past demography of the aurochs. a, Temporal trajectories of effective population size estimated using the PSMC applied to high-coverage genomes from each of the major aurochs ancestral strands. Three different trajectories are apparent: those of *Bos indicus*, North Asian wild and that shared among European aurochsen and Neolithic *Bos taurus*. **b**, Magnification of the mtDNA phylogeny (Fig. 3b) focusing on the two major *B. taurus* haplogroups, T and Q, indicating coalescences around the window of initial domestication at 11–10 ka. **c**, ROH profiles of high-coverage and imputed ancient genomes. Levels of ROH increase in post-LGM Europe and Southwest Asia and again after domestication in Neolithic cattle. The final group is ordered by *B. indicus* admixture proportion, ranging from 11% to 74%.

history[29]. When we test for admixture (*D*(water buffalo, Bovini sp.; *B. indicus, B. primigenius*)), we repeatedly see significant allele sharing of *B. indicus* with outgroup Bovini, consistent with such an ancestral admixture (Extended Data Fig. 5).

The two North Asian high-coverage *B. primigenius* genomes share similar PSMC profiles. However, from around 100 ka, these diverged from the trajectory shared by three European aurochsen and an early *B. taurus*; principally, by around 40 ka they had a markedly larger population size. During the LGM, there was a common sharp population decline, commensurate with habitat contraction that may have been exacerbated by the affinity of human populations for similar refugial biomes[30].

Forty-three (25 wild, 18 domesticated) of our ancient genomes have a coverage of more than 0.5×, and we leveraged these data, using imputation, to generate millions of genome-wide genotype calls[31]. This allowed an additional, sensitive window into past demography through the identification of individual profiles of runs of homozygosity (ROH)[32]. A clear regional pattern in the extent of ROH within genomes is observed. Similar fractions of the genome under short ROH are observed in North Asian *B. primigenius* sampled both before and after the LGM (Fig. 4c). By contrast, progressively higher levels of ROH in European aurochsen from before to after the LGM is consistent with a more restricted glacial population size in the west. *Bos indicus* and their hybrids show very low extents of ROH, consistent with a large population size.

Only three genomes possess a large fraction of long ROH within our wild cattle dataset, indicative of breeding between relatives, and two of these originate from islands (Sicily and North Jutlandic Island). This implies that aurochsen did not breed predominantly with close relatives; wild bovids often have male dispersal behaviours, which minimize inbreeding[33].

Domestication of the aurochs

The wild aurochs genomes documented here also clarify two key elements of the cattle domestication process: the initial capture of aurochsen by around 11 ka, and repeated secondary introgressions that further shaped domesticated *Bos* populations. The ROH profiles of seven Neolithic cattle show the highest levels of short (1–4 Mb) ROH segments, indicating a domestication-linked population restriction, but a lack of long ROH, suggesting inbreeding avoidance, probably through bull choice by herders.

The time-stamped mtDNA phylogeny (Fig. 3b) shows that considerable diversity existed in the wild, and that domestication captured only a limited subset of this genetic variation, with ancient and modern cattle sequences associating in tight, toothcomb-like clusters, each with recent common ancestors. Of note, more than 98% of sampled ancient and modern *B. taurus* mtDNA falls into just two such clades, labelled T and Q[34]. Within both of these clusters, a coalescence of just one or very few maternal ancestors is observed around the period of time, around 11 ka, when people first started experimenting with *Bos* husbandry (Fig. 4b). Given that aurochsen were some of the largest animals in ancient Southwest Asia, the initial capture and the controlled breeding of females are likely to have posed formidable challenges. Our results suggest that the capture and isolation of aurochsen were achieved on very few occasions, realized within a discrete historical window, and involved small numbers of individuals. This finding is congruent with human-led, directed *Bos* husbandry—including intentional foddering[35]—rather than a passive, gradual process[36,37]. In symmetry, the two main *B. indicus* clusters each coalesce around the time of Indus Valley domestication, around 8 ka.

After domestication, aurochsen and herded cattle coexisted for several millennia in many regions. From ancient cattle from Monjukli Depe (6.4 ka) in Turkmenistan) and Altai (around 4 ka) in Kazakhstan, we infer that introgression from the North Asian aurochs occurred at an early stage as cattle migrated eastwards (Fig. 2c and Extended Data Fig. 4). This process mirrors the introgression of the aurochs into cattle populations in Europe and the Levant that was previously inferred from whole-genome analyses[9,10,38]. Together, we infer that the mating of aurochsen with cattle was pervasive during the earliest spread of cattle herding across the diverse terrains and cultures of Eurasia. This admixture might have facilitated adaptation, and wild gene flow has been highlighted previously as part of a more diffuse view of domestication[39]. The introgression was male biased; our analyses show that there was more wild allele sharing in autosomes relative to the X chromosome in ancient domesticates, implying a reduced relative contribution from aurochs cows (Extended Data Fig. 6). This mirrors the later spread of *B. indicus* domestic ancestry beyond its South Asian region of origin, which is also inferred to have been male-mediated[10].

Wild and domestic cattle are not reciprocally monophyletic in a Y-chromosome-based phylogeny, and its topology suggests at least six separate origins within *B. taurus* (Extended Data Fig. 7). Introgression of an aurochs Y chromosome (Y1) has previously been suggested in European cattle[40], and we accordingly detect a Y1 haplotype in Neolithic Bulgarian aurochsen, which is an outgroup to the haplotypes observed in modern variation. Matrilineal contributions of introgressing aurochs are also detectable, but persist at a low frequency[41]. Divergent mtDNA clusters P and R (Fig. 3b) are inferred from their internal coalescences to have introgressed 3.7 ka and 1.9 ka, post-dating initial domestication by several thousand years.

Conclusions

Within ancient and modern *Bos* genomes we detect four major population ancestries, delineated by both autosomal and mtDNA variation. These each had demographic trajectories that were differentially

influenced by climate change and domestication. The most divergent are *B. indicus*, descended from *B. namadicus* in the tropical South Asian centre of Bovini diversity. Within *B. primigenius*, those from North Asia diverged from European and Southwest Asian populations during the last glacial period (around 90 ka). This North Asian ancestry persists in an unadmixed form in the Caucasus until 13.9 ka, but by 7 ka, genomes in that region are hybrids, perhaps reflecting the opening of habitat corridors after the advent of the Holocene. European wild genomes, despite sharing the common *B. primigenius* ancestor 100 ka, contain trace ancestry from much older, archaic European aurochsen. They have the strongest signal of glacial population constriction and also show a gradient of admixture with the neighbouring Southwest Asian type, which differentiates Iberian, Italian and Balkans refugial populations. Notably, much of Europe was recolonized from the former, although denser temporal sampling of European aurochsen during the last 50 ka might provide a higher resolution of refugial dynamics. The legacy of *B. primigenius* endures in domestic taurine cattle and this is inferred, from high levels of Neolithic ROH and a small number of initial matrilines, to have resulted from the restricted capture of Southwest Asian aurochsen. However, this diversity is augmented by early and pervasive male-mediated wild introgression, evident from contributions by each of the four separate preglacial aurochs ancestries that persist among the domestic cattle of today.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-08112-6.

1. van Vuure, C. & van Vuure, T. *Retracing the Aurochs: History, Morphology and Ecology of an Extinct Wild Ox* (Pensoft Publishers, 2005).
2. de Carvalho, C. N. et al. Aurochs roamed along the SW coast of Andalusia (Spain) during Late Pleistocene. *Sci. Rep.* **12**, 9911 (2022).
3. Wang, X., Flynn, L. J. & Fortelius, M. *Fossil Mammals of Asia: Neogene Biostratigraphy and Chronology* (Columbia Univ. Press, 2013).
4. Schulz, E. & KaiSer, T. M. Feeding strategy of the Urus *Bos primigenius* BOJANUS, 1827 from the Holocene of Denmark. *Cour. Forsch. Inst. Senckenberg* **259**, 155–164 (2007).
5. Bro-Jørgensen, M. H. et al. Ancient DNA analysis of Scandinavian medieval drinking horns and the horn of the last aurochs bull. *J. Archaeol. Sci.* **99**, 47–54 (2018).
6. Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proc. Natl Acad. Sci. USA* **115**, 6506–6511 (2018).
7. Bailey, J. F. et al. Ancient DNA suggests a recent expansion of European cattle from a diverse wild progenitor species. *Proc. Biol. Sci.* **263**, 1467–1473 (1996).
8. Troy, C. S. et al. Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410**, 1088–1091 (2001).
9. Park, S. D. E. et al. Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle. *Genome Biol.* **16**, 234 (2015).
10. Verdugo, M. P. et al. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science* **365**, 173–176 (2019).
11. Ginja, C. et al. Iron age genomic data from Althiburos – Tunisia renew the debate on the origins of African taurine cattle. *iScience* **26**, 107196 (2023).
12. Svendsen, J. I. et al. Late Quaternary ice sheet history of northern Eurasia. *Quat. Sci. Rev.* **23**, 1229–1271 (2004).
13. Wright, E. in *Cattle and People: Interdisciplinary Approaches to an Ancient Relationship* (eds Wright, E. & Ginja, C.) 3–27 (Lockwood Press, 2022).
14. Hewitt, G. Post-glacial re-colonization of European biota. *Biol. J. Linn. Soc. Lond.* **68**, 87–112 (1999).
15. Linseele, V. Size and size change of the African aurochs during the Pleistocene and Holocene. *J. Afr. Archaeol.* **2**, 165–185 (2004).
16. Lefèvre, D., Raynal, J.-P., Vernet, G., Kieffer, G. & Piperno, M. Tephro-stratigraphy and the age of ancient Southern Italian Acheulean settlements: the sites of Loreto and Notarchirico (Venosa, Basilicata, Italy). *Quat. Int.* **223–224**, 360–368 (2010).
17. Pereira, A. et al. The earliest securely dated hominin fossil in Italy and evidence of Acheulian occupation during glacial MIS 16 at Notarchirico (Venosa, Basilicata, Italy). *J. Quat. Sci.* **30**, 639–650 (2015).
18. Cassoli, P. F., Di Stefano, G. & A., T. in *Notarchirico: Un Sito del Pleistocene Medio Iniziale nel Bacino* (ed. di Piperno, M.) 361–438 (Osanna, 1999).
19. Lambeck, K., Esat, T. M. & Potter, E.-K. Links between climate and sea levels for the past three million years. *Nature* **419**, 199–206 (2002).
20. Batchelor, C. L. et al. The configuration of Northern Hemisphere ice sheets through the Quaternary. *Nat. Commun.* **10**, 3713 (2019).
21. Otvos, E. G. The Last Interglacial Stage: definitions and marine highstand, North America and Eurasia. *Quat. Int.* **383**, 158–173 (2015).
22. Leonardi, M., Boschin, F., Boscato, P. & Manica, A. Following the niche: the differential impact of the last glacial maximum on four European ungulates. *Commun. Biol.* **5**, 1038 (2022).
23. Helmens, K. F. The Last Interglacial–Glacial cycle (MIS 5–2) re-examined based on long proxy records from central and northern Europe. *Quat. Sci. Rev.* **86**, 115–143 (2014).
24. Kosintsev, P. A. & Bachura, O. P. Late Pleistocene and Holocene mammal fauna of the Southern Urals. *Quat. Int.* **284**, 161–170 (2013).
25. Robin, M. et al. Ancient mitochondrial and modern whole genomes unravel massive genetic diversity loss during near extinction of Alpine ibex. *Mol. Ecol.* **31**, 3548–3565 (2022).
26. Miller, W. et al. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proc. Natl Acad. Sci. USA* **109**, E2382–E2390 (2012).
27. Lord, E. et al. Population dynamics and demographic history of Eurasian collared lemmings. *BMC Ecol. Evol.* **22**, 126 (2022).
28. Murray, C., Huerta-Sanchez, E., Casey, F. & Bradley, D. G. Cattle demographic history modelled from autosomal sequence variation. *Phil. Trans. R. Soc. B* **365**, 2531–2539 (2010).
29. Sinding, M.-H. S. et al. Kouprey (*Bos sauveli*) genomes unveil polytomic origin of wild Asian Bos. *iScience* **24**, 103226 (2021).
30. Bergman, J. et al. Worldwide Late Pleistocene and Early Holocene population declines in extant megafauna are associated with *Homo sapiens* expansion rather than climate change. *Nat. Commun.* **14**, 7679 (2023).
31. Erven, J. A. M. et al. A high-coverage Mesolithic aurochs genome and effective leveraging of ancient cattle genomes using whole genome imputation. *Mol. Biol. Evol.* **41**, msae076 (2024).
32. MacLeod, I. M., Larkin, D. M., Lewin, H. A., Hayes, B. J. & Goddard, M. E. Inferring demography from runs of homozygosity in whole-genome sequence, with correction for sequence errors. *Mol. Biol. Evol.* **30**, 2209–2223 (2013).
33. Clutton-Brock, T. *Mammal Societies* (Wiley, 2016).
34. Cubric-Curik, V. et al. Large-scale mitogenome sequencing reveals consecutive expansions of domestic taurine cattle and supports sporadic aurochs introgression. *Evol. Appl.* **15**, 663–678 (2022).
35. Makarewicz, C. & Tuross, N. Finding fodder and tracking transhumance: isotopic detection of goat domestication processes in the Near East. *Curr. Anthropol.* **53**, 495–505 (2012).
36. Bollongino, R. et al. Modern taurine cattle descended from small number of near-eastern founders. *Mol. Biol. Evol.* **29**, 2101–2104 (2012).
37. Scheu, A. et al. The genetic prehistory of domesticated cattle from their origin to the spread across Europe. *BMC Genet.* **16**, 54 (2015).
38. Upadhyay, M. R. et al. Genetic origin, admixture and population history of aurochs (*Bos primigenius*) and primitive European cattle. *Heredity* **118**, 169–176 (2017).
39. Marshall, F. B., Dobney, K., Denham, T. & Capriles, J. M. Evaluating the roles of directed breeding and gene flow in animal domestication. *Proc. Natl Acad. Sci. USA* **111**, 6153–6158 (2014).
40. Götherström, A. et al. Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. *Proc. Biol. Sci.* **272**, 2345–2350 (2005).
41. Achilli, A. et al. The multifaceted origin of taurine cattle reflected by the mitochondrial genome. *PLoS One* **4**, e5753 (2009).

[1]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland. [2]Department of Biology, University of Copenhagen, Copenhagen, Denmark. [3]Palaeogenetics Group, Institute of Organismic and Molecular Evolution (iomE), Johannes Gutenberg University Mainz, Mainz, Germany. [4]Groningen Institute of Archaeology, University of Groningen, Groningen, The Netherlands. [5]School of Agriculture and Food Science, University College Dublin, Dublin, Ireland. [6]Globe Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [7]Bioinformatics Support Unit, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK. [8]Palaeogenomics and Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford, UK. [9]McDonald Institute for Archaeological Research, University of Cambridge, Cambridge, UK. [10]Independent researcher, Langcliffe, UK. [11]Institute of Archaeology and Steppe Civilizations, Al-Farabi Kazakh National University, Almaty, Kazakhstan. [12]Sezione di Geologia e Paleontologia, Museo Civico di Storia Naturale di Verona, Verona, Italy. [13]Vendsyssel Historical Museum, Hjørring, Denmark. [14]Department of Natural Sciences, National Museums Scotland, Edinburgh, UK. [15]School of Geosciences, University of Edinburgh, Edinburgh, UK. [16]LEIZA, Archaeological Research Centre and Museum for Human Behavioural Evolution, Schloss Monrepos, Neuwied, Germany. [17]Palaeogenomics Group, Institute of Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, Ludwig-Maximilians-Universität, Munich, Germany. [18]School of Biological and Chemical

Sciences, Queen Mary University of London, London, UK. [19]Institute for Prehistory and Protohistory, University of Cologne, Cologne, Germany. [20]Bioarchaeology Service, Museo delle Civiltà, Piazza Guglielmo Marconi, Rome, Italy. [21]Sezione di Geologia e Paleontologia, Museo della Natura e dell' Uomo, Padova, Italy. [22]Dipartimento di Scienze della Terra, Università di Pisa, Pisa, Italy. [23]Department of Paleontology and Mineralogy, National Museum of Natural History, Bulgarian Academy of Sciences, Sofia, Bulgaria. [24]National Museum of Denmark, Copenhagen, Denmark. [25]Department of Recreational Geography, Service, Tourism and Hospitality, Institute of Geography, Altai State University, Barnaul, Russian Federation. [26]The Biological Museum, Lund University, Arkivcentrum Syd, Lund, Sweden. [27]Instituto Universitario de Xeoloxía, Universidade da Coruña (UDC), A Coruña, Spain. [28]German Archaeological Institute, Central Department, Berlin, Germany. [29]Lauresham Laboratory for Experimental Archaeology, UNESCO-Welterbestätte Kloster Lorsch, Lorsch, Germany. [30]Department of Archaeology, Ethnography and Museology, Altai State University, Barnaul, Russian Federation. [31]Toraighyrov University, Joint Research Center for Archeological Studies, Pavlodar, Kazakhstan. [32]Department of Archaeology and History, University of Exeter, Exeter, UK. [33]Department of Archaeology and Ancient History, Lund University, Lund, Sweden. [34]Centre for Palaeogenetics, Stockholm, Sweden. [35]Lund University Historical Museum, Lund, Sweden. [36]Paleoecology Laboratory, Institute of Plant and Animal Ecology, Ural Branch of the Russian Academy of Sciences, Ekaterinburg, Russian Federation. [37]Department of History, Institute of Humanities, Ural Federal University, Ekaterinburg, Russian Federation. [38]Zoological Institute of the Russian Academy of Sciences, Saint Petersburg, Russian Federation. [39]Archaeology Stable Isotope Laboratory, Institute of Pre- and Protohistoric Archaeology, University of Kiel, Kiel, Germany. [40]University of Haifa, Haifa, Israel. [41]Deceased: Viktor Zeibert. ✉e-mail: mhssinding@gmail.com; dbradley@tcd.ie

Methods

DNA extraction and sequencing of archaeological samples

Ancient bone and teeth samples were processed in designated ancient DNA facilities. Most samples were processed at Trinity College Dublin, with additional processing in the ancient DNA facilities of the Palaeogenetics Group at Johannes Gutenberg University Mainz, the PalaeoBARN laboratory at the University of Oxford and the Globe Institute at the University of Copenhagen.

Trinity College Dublin. Before drilling, bones and teeth were exposed to UV radiation for 30 min on each side. A dentist drill was then used to clean the surface and the exposed element was then irradiated for a further 15 min. A saw (Dremel) was used to remove bone, which was then pulverized to a fine powder using a Mixer Mill (MM400, Retsch). A 100–135-mg sample of the powdered bone was used for DNA extraction.

The samples Baikal1, Gyu1, Tango1 and Tango2 were processed according to a previous protocol[42], with fully described modifications[43]. The remaining samples were processed using an extraction method involving a dilute (0.5%) sodium hypochlorite pre-wash based on previous protocols[44,45] and fully described[46]. Purified DNA (16.25 µl) was treated with 5 µl of (1 U/1 µl) USER enzyme (New England BioLabs) and incubated for 3 h at 37 °C, followed by Blunt-End dsDNA library preparation[47], modified as per a previous study[48]. AccuPrime Pfx DNA Polymerase (Invitrogen) was used for the PCR amplification with 3 µl of DNA library and PCR cycles varying from 10–16 cycles. PCR products were purified using minElute columns (Qiagen) and were quantified using a Tapestation 2200 (Agilent Technologies).

Indexed PCR libraries were sequenced on a HiSeq 2500 (SE 100 bp) at Macrogen (Seoul) and a NovaSeq 6000 (PE 150 bp) at TrinSeq (Dublin).

Johannes Gutenberg University. DNA extractions were performed according to previously detailed methods[10,49], with slight modifications to the extraction protocol. Extracts underwent UDG treatment and subsequently double-stranded sequencing libraries were prepared and sequenced (100 bp SE) on an Illumina Hiseq 2000 at Macrogen (Seoul).

Oxford University. Approximately 200 mg bone was sampled from the petrous part of the temporal bone using an electric saw. DNA extraction was performed on 112.8 mg of powder following a previous protocol[44]. A double-stranded library was built using the Blunt-End Single Tube protocol[50] and, after this step, the amount of DNA was estimated using a quantitative PCR assay to adjust for the number of PCR cycles to use in library amplification. Indexed PCR libraries were sequenced on a NovaSeq 6000 (PE 150 bp) at TrinSeq (Dublin).

Ancient data processing

Single-end sequencing reads were trimmed using cutadapt v.3.5[51] (cutadapt -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -O 1 -m 30). Paired-end sequencing reads were trimmed and collapsed using AdapterRemoval v.2.3.0[52] (AdapterRemoval --file1 reads_1.fq.gz --file2 reads_2.fq --basename output_paired --trimns --trimqualities --minquality 25 --collapse). The collapsed fastq file was used in downstream analysis and the truncated read pairs were not used.

All trimmed reads of samples were aligned to ARS-UCD1.2 (edited to include a bovine Y chromosome, GCF_000003205.7) using BWA[53] v.0.7.5a with relaxed parameters (-l 1024 -n 0.01 -o 2). The resulting SAI files were converted to SAM file format using BWA samse[53]. SAM files were then converted into binary BAM files using SAMtools (v.1.13) view[54]. The -F4 flag was used with SAMtools view to discard unmapped reads for all BAM files. The BAM files were then sorted using SAMtools sort. Read groups were added to each BAM file using AddOrReplaceReadGroups (https://broadinstitute.github.io/picard/), including information on index and library preparation. Picard (v.2.22.1) was then used to remove optical and PCR duplicates using MarkDuplicates (REMOVE_DUPLICATES=true). Indels were locally realigned using RealignerTargetCreator from GATK (v.2.4). BAM files were then filtered for reads of a mapping quality less than 25 using SAMtools view.

BAM files were assessed using mapDamage2.0[55]. mapDamage2.0 calculates read length distributions and substitution patterns at the 5′ and 3′ ends of aligned read BAM files. For USER-treated BAM files, PMDTools was also used to calculate deamination rates at CpG sites only, which cannot be excised by USER treatment[56] (--deamination --range 30 --CpG). Depth of coverage was calculated using Qualimap v.2.1.3[57]. All results are detailed in Supplementary Data 1 and 2.

Modern data processing

Raw paired-end sequencing data for 290 modern cattle were downloaded from the European Nucleotide Archive (ENA). The read pairs were trimmed using Trimmomatic (v.0.39)[58], then aligned to the reference genome using bwa mem[59]. Paired read mate coordinates and insert size were added with SAMtools fixmate, and sorted using SAMtools sort. Read groups were added to each BAM file using Picard (v.2.22.1) AddOrReplaceReadGroups (https://broadinstitute.github.io/picard/), and Picard (v.2.22.1) was then used to remove optical and PCR duplicates using MarkDuplicates (REMOVE_DUPLICATES=true). Indels were locally realigned using RealignerTargetCreator from GATK (v.2.4). BAM files were then filtered for reads of a mapping quality less than 25 using SAMtools view. Depth of coverage was calculated using Qualimap v.2.1.3[57].

In addition, raw paired-end sequencing data for 37 outgroup Bovinis were downloaded from ENA and processed, as above. All results are detailed in Supplementary Data 3.

Diploid calling

A total of 243 ancient and modern cattle and aurochs genomes with a mean coverage of more than 8× were called together. BCFtools (v.1.12) mpileup and call were used to call autosomal single-nucleotide polymorphisms (SNPs) from this dataset (--skip-indels -q 30 -C 50 -Q 20 -B -a FORMAT/AD, FORMAT/SP, FORMAT/ADF, FORMAT/ADR, FORMAT/DP, INFO/AD) and generate VCF files (call -mv -f GQ,GP)[60]. Individual calls with coverages of one-third or thrice the mean coverage of the sample were marked as missing ("./."). An allelic balance filter of 0.2 was applied to each individual call using Picard (v.2.22.1), setting any variant with an allelic balance between 0.1 and 0.2 to missing (FilterVcf --MIN_AB 0.2). Variants with a quality score of less than 20 were filtered using BCFtools view. Tri-allelic, quad-allelic and SNPs within 3 bp of variants were removed using VCFtools v.0.0.17 (--thin 3 --min-alleles 2 --max-alleles 2). CpG sites were removed and a site missingness filter of 20% was applied (--geno 0.2), and individuals with a missingness of 20% or more were removed (--mind 0.2) using VCFtools[61]. This resulted in a total of 45,623,135 biallelic SNPs. All SNPs present in repeat regions, as defined by the UCSC Browser and Repeat-Masker files (https://genome.ucsc.edu/), were removed. Subsequently, singleton heterozygous or singleton homozygous variants were removed. This left a total of 21,672,432 biallelic SNPs. SNPs were also called on the X chromosome, as above, except that the --ploidy 1 flag was also applied when calling SNPs in males with BCFtools call. This resulted in a total of 773,299 SNPs on the X chromosome after applying all filtering steps as above.

Separately, autosomal SNPs were called in 24 outgroup Bovini species above a mean coverage of 8×. The fully filtered SNP dataset resulted in a total of 14,510,692 sites.

Pseudohaploid calling

For population genomic analysis, three transversion SNP datasets were prepared. First, the transition SNPs from the 21,672,432 biallelic autosomal SNP set were filtered, leaving a total of 7,108,856 sites. Second, the transition SNPs of the 773,299 biallelic X chromosomal SNP set were filtered, leaving a total of 255,371 sites. Third, the transition SNPs of

the 14,510,692 biallelic outgroup Bovini SNP set were filtered, leaving a total of 4,732,559 sites.

Pseudohaploid SNP calls were using ANGSD (v.0.941-11-g7a5e0db) Haplocaller by sampling a random base called using the SAMtools genotype likelihood algorithm (-doHaploCall 1 -doCounts 1 -GL 1) with several filtering steps for quality control (-C50 -uniqueOnly 1 -remove_bads 1 -minMapQ 30 -minQ 25)[62]. USER-treated ancient samples were called with the -trim 4 command and non-USER-treated ancient samples were called with -trim 6. The three SNP datasets were called separately on a total of 455 samples (290 modern *Bos* genomes, 128 ancient *Bos* genomes and 37 outgroup genomes; Supplementary Data 1–3) by supplying the SNP coordinates using the -sites flag.

mtDNA alignment

Mitochondrial reference genome assemblies for each major haplogroup were constructed using a 'baiting and iterative mapping' approach based on the methods in a previous study[63]. Because mitochondrial genomes are circular, sequencing reads that straddle the artificial start or end-point of the genome assembly may be lost. To this end, each reference FASTA was circularized by concatenating the last 15 bp of the FASTA reference to the start of the reference and concatenating the first 15 bp to the end of the reference FASTA.

For each sample, all sequencing data were aligned to their haplotypic reference genome and processed as above, without indel realignment. A consensus sequence was generated using ANGSD (v.0.941-11-g7a5e0db) (angsd -doFasta 2 -doCounts 1 -setMinDepth 3 -minQ 20 -minMapQ 30) and the resulting FASTA was decircularized by removing the first and last 15 bp of the sequence. Indels were identified using Integrative Genomics Viewer (1.13.11)[64], and manually corrected in the FASTA file.

BEAST

Bayesian analyses of mitochondrial sequences were performed using BEAST v.2.6.6[65]. Partitioning of data was done on the basis of the NCBI partition annotations of the *B. taurus* mitochondrial reference genome (V00654.1). The partitions were tRNA, rRNA, the first and second codon positions of gene regions (C1 + 2), the third codon position (C3), the D-loop and the remainder of the molecule. The sequence data (minimum 10× coverage) were prepared by performing a multiple sequence alignment using MUSCLE[66], converted to a Nexus file using SeaView v.5.0.1[67], and loaded into BEAUTi. For ancient samples, tip dates were set to a uniform probability distribution of the radiocarbon 95% confidence interval. The site model of each partition was chosen by bModelTest 1.2.1[68], and the clock model was set to Strict. A coalescent Bayesian skyline model was chosen as the tree model[69], and a BEAST xml file was generated.

Four independent BEAST runs of 250 million chains were performed with a random seed. The MCMC chain was sampled every 25,000 iterations, with 10% burn-in. Log files were assessed using Tracer v.1.7.2 to inspect parameter convergence[70]. All independent BEAST runs converged after 250 million chains, with all parameters reaching an effective sampling size of more than 300. They were combined using LogCombiner and a maximum clade credibility tree was constructed with median node heights using TreeAnnotator. The resulting tree was visualized with FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Y chromosome

The bovine male-specific region of the Y chromosome consists of three major regions: the Y-transitional, the Y-ampliconic and the X-degenerate[71]. The X-degenerate region spans 2.5–3.9 Mb and 42.2–43.3 Mb along the bovine Y assembly, GCF_000003205.7. A mappability filter based on H. Li's SNPable filter (http://lh3lh3.users.sourceforge.net/snpable.shtml) was applied to this 2.5-Mb region, reducing the total usable region to 1,443,572 total bases.

For variant discovery, BCFtools (v.1.12) mpileup and call were used to call SNPs from all male *Bos* genomes above an average coverage of 1× (--skip-indels -q 30 -C 50 -Q 20 -B -a FORMAT/AD, FORMAT/SP, FORMAT/ADF, FORMAT/ADR, FORMAT/DP and INFO/AD), and used to generate VCF files (call -mv -f GQ,GP --ploidy 1). Individual calls with coverages of a minimum of 2 and less than one-third or more than thrice the mean coverage of the sample were marked as missing ("./."). Variants with a quality score of less than 20 were filtered using BCFtools view. Tri-allelic, quad-allelic and SNPs within 3 bp of variants were removed using VCFtools (--thin 3 --min-alleles 2 --max-alleles 2). CpG sites were removed. A missingness filter of 20% was applied using VCFtools (--geno 0.2), leaving a total of 4,943 biallelic sites. These sites were called for all males using BCFtools (v.1.12) mpileup, as above. The VCF file was converted to a multifasta file and samples with a missingness of more than 15% were removed. The multifasta file was loaded into SeaView v.5.0.1[67] and a maximum likelihood tree was constructed with PhyML with 100 bootstraps[72].

Imputation

To prepare our reference dataset, the 21,672,432-site VCF file was phased using Beagle 5.2 (v.28Jun21.220) with no imputation (impute=false) and the effective population size was set to 20,000 (ne=20000)[73,74]. Using this SNP set, genotype likelihood calls of all ancient samples with more than 0.5× coverage were generated using BCFtools (v.1.12) mpileup (-I,-E, -a "FORMAT/DP, FORMAT/AD, INFO/AD") and call (-Aim -C alleles -T). Imputation was performed on the ancient genotype likelihood calls separately using GLIMPSE v.1.1.1[75], following the GLIMPSE pipeline. The imputed data were filtered for a strict genotype probability (GP) ≥ 0.99 and an INFO score ≥ 0.99 and were merged with the diploid called dataset.

To test the accuracy of imputation, the nine high-coverage ancient samples were downsampled to 0.5× and 1.0× on a chromosomal level using Picard v.2.20.0. The high-coverage sample was then removed from the phased VCF dataset and the downsampled sample was imputed with this filtered reference panel. Imputation accuracy was calculated with Picard's GenotypeConcordance tool. The high-coverage genotypes were used as the TRUTH_VCF, and the imputed genotypes were used as the CALL_VCF.

ROH analysis

A dataset of 58 animals was made to include modern taurines (less than 1% missingness), modern indicine (less than 5% missingness), ancient domesticates (less than 10% missingness) and all aurochs. Using PLINK (v.1.9), a genotyping filter of 0 (-geno 0) and a minor allele frequency (MAF) filter of 1% (-maf 0.01) were applied, leaving 2,948,566 sites. ROH were estimated with PLINK (--homozyg --homozyg-density 50 --homozyg-gap 100 --homozyg-kb 500 --homozyg-snp 50 --homozyg-window-het 1 --homozyg-window-snp 50 --homozyg-window-threshold 0.05).

MSMC2

MSMC2 (v.2.1.3) was used to produce PSMC' curves for all ancient samples above an average genome coverage of 8× and a selection of modern genomes[76]. For individual PSMC' curves, 45,623,135 biallelic SNPs were used on each unphased genome. A mappability filter based on H. Li's SNPable filter was also produced for each chromosome, filtering for sites that scored highly for unique mapping (score of 3) (http://lh3lh3.users.sourceforge.net/snpable.shtml). Coverage filters were generated using bamCaller.py from the msmc-tools repository for each individual genome. A negative mask was also supplied for CpG sites. MSMC2 was run on each genome with 20 bootstraps and the time segment pattern: --timeSegmentPattern 25*1 + 1*2 + 1*3. Curves were scaled with a mutation rate of 1.26×10^{-8} and a generation time of 6, as described previously[77].

MSMC2 was also used to perform cross-population analysis to determine the approximate divergence time between populations. Up to three samples were extracted from the phased 21,672,432 biallelic SNP

dataset to represent each population (European wild, North Asian, Near Eastern and indicine). MSMC2 was run on each to estimate coalescence rates within each population, supplying mask filters, as above. An MSMC2 run was then done on all cross-population pairs to estimate the cross-coalescence rate, again supplying mask filters. Finally, the combineCrossCoal.py script from the msmc-tools was used to plot the relative cross-coalescence rate.

Pseudodiploid X chromosome analysis was also performed on pairs of male genomes. BCFtools (v.1.12) mpileup and call were used to call SNPs on the haploid male X chromosome as above, except emitting all non-variant and uncalled sites (call -m -f GQ,GP --ploidy 1). A mask file containing sites with coverages less than half or more than twice the mean X chromosome coverage was produced for each sample using bamCaller.py in the msmc-tools package. Pairs of male X chromosomes were then combined into one 'pseudodiploid' VCF file. Using MSMC2, PSMC' was performed on each haploid X pair, supplying a mask file to exclude sites that were above or below the coverage threshold in either sample, a mappability mask based on H. Li's SNPable filter (http://lh3lh3.users.sourceforge.net/snpable.shtml) and CpG sites. The mutation rate was scaled down by 25% to account for the lower mutation rate of the X chromosome relative to the autosome.

Identity by state

To compare patterns of allele sharing between genomes, pairwise identity-by-state (IBS) matrices were computed on the pseudohaploid dataset using Plink v.1.9 (-distance 1-ibs)[78]. A classical MDS of the resulting data matrix was computed in R v.4.2.2. An IBS matrix was also computed with a water buffalo sample and a neighbour joining tree was produced using ape v.5.6.2[79], setting water buffalo as the outgroup. The tree was visualized using FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Principal component analysis

PCAngsd (v.1.01)[80] was used to produce principal component analyses (PCAs) on several *Bos* datasets. PCAs were computed using pseudohaploid genotypes of ancient and modern genomes with ≤0.95 missingness. To account for linkage disequilibrium (LD), the SNP dataset was pruned in PLINK (v.1.9) with the command --indep-pairwise 50 5 0.5, leaving a total of 5,474,894 variants. Each dataset was filtered and a MAF filter of 2.5% was applied prior to computing the PCA.

D statistics

To calculate excess allele sharing, test for treeness, and determine the evidence gene flow, *D* statistics[81] were computed on pseudohaploid datasets with a 2.5% MAF filter using ADMIXTOOLS 2[82]. Water buffalo was set as the outgroup.

NGSadmix

NGSadmix (v.32) is a tool for estimating individual admixture proportions of low-coverage data by using genotype likelihoods, incorporating the uncertainty introduced by unobserved genotypes[83]. Genotype likelihoods were computed using ANGSD (v.0.941-11-g7a5e0db) for a dataset of 165 *Bos* samples comprising one modern animal per breed and all ancient samples with ≤0.95 missingness. Genotype likelihoods were called using the GATK model (-GL 2) supplying a site file of the LD-pruned 5,474,894 transversion-only variants (-domajorminor 1 -sites) and the file was dumped to a Beagle likelihood file (-doGlf 2). A genotype filter of 75% and a MAF filter of 2.5% were applied, leaving a total of 1,928,771 SNPs.

NGSadmix was run using a range of *K* values (1–10) and the likelihood value from each run of NGSadmix was extracted and written to a logfile. This process was repeated 50 times and this file was used with Clumpak v.1.1[84] to calculate the most likely *K* on the basis of methods described previously[85].

Treemix

Treemix (v.1.13)[86] was used to model *Bos* populations as a phylogeny of population splits, allowing for gene flow between branches in the form of migration edges. Ancient and modern populations were extracted from the MAF 2.5% pseudohaploid call set and water buffalo was included as an outgroup. SNP sites were filtered out using PLINK (v.1.9) if not present in at least one individual in each group.

Treemix was run using windows of 1,000 SNPs, and using the --noss and --global parameters to disable sample-size correction and allow for global rearrangements, setting water buffalo as the outgroup. Residuals were examined for the models at each *m*.

AdmixtureGraph

To further explore models of *Bos* populations as a phylogeny, admixture graphs were fitted using ADMIXTOOLS 2[82]. Using the MAF 2.5% pseudohaploid call set, f_2 values were extracted for pairs of populations using extract_f2(maxmiss = 1, auto_only=F). To determine the optimal number of migration events, find_graphs() was run for $m = 0–8$ with no constraints, setting water buffalo as the outgroup and setting stop_gen = 100. This was repeated 100 times and the worst residual and the log-likelihood of each best-fit model were recorded and the distribution of both statistics were assessed for each *m* to determine the least migration events which may explain the relationships of the populations. Find_graphs() was run for 250 iterations and the graphs were manually evaluated. Feasible graph fits were compared with qpgraph_resample_multi and compare_fits functions, with 100 resampled SNP block set bootstraps to test whether the likelihood score of a chosen graph was significantly higher than that of the other graph(s).

qpAdm

To test for multiple ancestries in *B. primigenius* samples, models were tested with qpAdm[87,88] using ADMIXTOOLS 2[82]. Using the MAF 2.5% pseudohaploid call set, qpAdm was used with auto_only=FALSE and allsnps=TRUE.

Sliding-windows F_{ST}

Sliding F_{ST} was calculated between modern taurine breeds and aurochs to investigate regions in the genome that are strongly differentiated between wild and domestic populations. Twelve postglacial European aurochs with an average coverage of more than 2× were included in the wild population, and a balanced selection of taurine breeds (four European breeds, four African breeds and four Asian breeds) were included in the domestic population. Using ANGSD (v.0.941-11-g7a5e0db), the site frequency spectrum was calculated with sites covered in at least half of the individuals in the group with a minimum depth of 2 in each individual (-minInd 6 -minIndDepth 2). Only biallelic traversion sites were considered (-rmTriallelic 1e-4 -noTrans 1) with a minimum base quality of 25 and a mapping quality of 30 (-minQ 25 -minMapQ 30). In addition, four bases of both ends of the read (-trim 4) were trimmed and only unique reads without excessive mismatches were considered (-uniqueOnly 1, -remove_bads 1 -C 50). Water Buffalo was used as an ancestral population (-anc). F_{ST} was calculated from the site frequency spectrums in 50 kb sliding windows with 10 kb steps. The top 0.1% of the F_{ST} windows were considered outliers and adjacent windows were combined to form F_{ST} outlier regions. The outlier regions and regions within 100 kb of the outlier regions were searched for genes using the RefSeq annotation NCBI Annotation Release 106.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Sequence reads and alignment files of new data are available through the ENA under accession number PRJEB75467. Previously published ancient data used in this study are available under accession numbers PRJEB31621, PRJNA379859, PRJNA803479 and PRJEB74338, and are detailed in Supplementary Table 2. Sources for previously published modern genomic data analysed here are listed in Supplementary Table 3. The ARS-UCD1.2 reference genome is available under NCBI assembly accession GCF_002263795.1. The *B. taurus* mitochondrial reference genome is available under NCBI accession number V00654.1.

Code availability

This study makes use of publicly available software, referenced throughout the main text and supplementary material.

42. Yang, D. Y., Eng, B., Waye, J. S., Dudar, J. C. & Saunders, S. R.Improved DNA extraction from ancient bones using silica-based spin columns. *Am. J. Phys. Anthropol.* **105**, 539–543 (1998).
43. Mattiangeli, V., Cassidy, L. M., Daly, K. G., Mullin, V. E. & Verdugo, M. Multi-step ancient DNA extraction protocol for bone and teeth. Protocols.io https://doi.org/10.17504/protocols.io.6qpvr45b2gmk/v1 (2023).
44. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
45. Boessenkool, S. et al. Combining bleach and mild predigestion improves ancient DNA recovery from bones. *Mol. Ecol. Resour.* **17**, 742–751 (2017).
46. Mattiangeli, V., Cassidy, L. M., Daly, K. G. & Mullin, V. E. Bleach extraction protocol: damaged or degraded DNA recovery from bone or tooth powder. Protocols.io https://doi.org/10.17504/protocols.io.8epv5j88nl1b/v1 (2023).
47. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, pdb.prot5448 (2010).
48. Gamba, C. et al. Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014).
49. Botigué, L. R. et al. Ancient European dog genomes reveal continuity since the Early Neolithic. *Nat. Commun.* **8**, 16082 (2017).
50. Carøe, C. et al. Single-tube library preparation for degraded DNA. *Methods Ecol. Evol.* **9**, 410–419 (2018).
51. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
52. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
53. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
54. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
55. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
56. Skoglund, P. et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl Acad. Sci. USA* **111**, 2229–2234 (2014).
57. Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294 (2016).
58. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
59. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1303.3997 (2013).
60. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
61. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
62. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
63. Hahn, C., Bachmann, L. & Chevreux, B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res.* **41**, e129 (2013).
64. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
65. Bouckaert, R. et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).
66. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
67. Gouy, M., Tannier, E., Comte, N. & Parsons, D. P. Seaview Version 5: a multiplatform software for multiple sequence alignment, molecular phylogenetic analyses, and tree reconciliation. *Methods Mol. Biol.* **2231**, 241–260 (2021).
68. Bouckaert, R. R. & Drummond, A. J. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evol. Biol.* **17**, 42 (2017).
69. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
70. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
71. Chang, T.-C., Yang, Y., Retzel, E. F. & Liu, W.-S. Male-specific region of the bovine Y chromosome is gene rich with a high transcriptomic activity in testis development. *Proc. Natl Acad. Sci. USA* **110**, 12373–12378 (2013).
72. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
73. Browning, B. L., Tian, X., Zhou, Y. & Browning, S. R. Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* **108**, 1880–1890 (2021).
74. Browning, B. L., Zhou, Y. & Browning, S. R. A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018).
75. Rubinacci, S., Ribeiro, D. M., Hofmeister, R. J. & Delaneau, O. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.* **53**, 120–126 (2021).
76. Schiffels, S. & Wang, K. MSMC and MSMC2: the Multiple Sequentially Markovian Coalescent. *Methods Mol. Biol.* **2090**, 147–166 (2020).
77. Chen, N. et al. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat. Commun.* **9**, 2337 (2018).
78. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
79. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
80. Meisner, J. & Albrechtsen, A. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719–731 (2018).
81. Soraggi, S., Wiuf, C. & Albrechtsen, A. Powerful inference with the D-statistic on low-coverage whole-genome data. *G3* **8**, 551–566 (2018).
82. Maier, R. et al. On the limits of fitting complex models of population history to f-statistics. *eLife* **12**, e85492 (2023).
83. Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).
84. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15**, 1179–1191 (2015).
85. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
86. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
87. Harney, É., Patterson, N., Reich, D. & Wakeley, J. Assessing the performance of qpAdm: a statistical tool for studying population admixture. *Genetics* **217**, iyaa045 (2021).
88. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
89. Lisiecki, L. E. & Raymo, M. E. A Pliocene–Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. *Paleoceanogr. Paleoclimatol.* **20**, PA1003 (2005).

Acknowledgements This work is funded by the European Research Council under the European Union's Horizon 2020 research and innovation program, grant agreements 885729-AncestralWeave and 295729-CodeX. M.-H.S.S. is supported by the Carlsberg Foundation (Reintegration Fellowship, CF20-0355) and the Independent Research Fund Denmark (International Postdoc, 8020-00005B), which also contributed to the sequencing costs and carbon dating costs of this project; E.R. is supported by the Jan Löfqvist Endowment and Axel and Margaret Ax:son Johnson Foundation for Public Benefit and contributed to the sequencing and carbon dating costs of this project; K.G.D. is supported by Science Foundation Ireland (grant number 21/PATH-S/9515/); V.E.M. is supported by a Government of Ireland Postdoctoral Fellowship (GOIPD/2020/605); A.G.-d. and A.V.-G. are supported by Xunta de Galicia (CN 2021/17); M. Sablin is supported by state assignment no. 122031100282-2; A.A.T. is supported by the Russian Science Foundation ('The World of Ancient Nomads of Inner Asia: Interdisciplinary Studies of Material Culture, Sculptures and Economy', project no. 22-18-00470); and C.A.M. is supported by the European Council for Research for a Horizon 2020 grant (ASIAPAST, action no. 772957). We thank TrinSeq for sequencing support; L. Cassidy and L. Wright for discussions; R. Verdugo for contributions to figure design; the Margulan Institute of Archaeology Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan for storing, archiving and providing the material from the Roschinskoe site; the Museum of the Institute of Plant and Animal Ecology (Ural Branch of the Russian Academy of Sciences, Ekaterinburg) for providing specimens for sampling; A. Zeeb-Lanz and R. Arbogast for permitting the sampling of bones from Herxheim; R. W. Schmitz for permitting the sampling of bones from Bedburg-Königshoven; and H. Hartnagel for assistance in sampling the Rhine specimens.

Author contributions D.G.B. and M.-H.S.S. conceived the project. Laboratory work was done by C.R., M.-H.S.S., V.E.M., A.S., M.P.V., K.G.D., M.M.C., V.M., D.D. A.M. and P.B. C.R. processed and analysed the data with contributions from M.-H.S.S., V.E.M., J.A.M.E. and K.G.D. All other authors gave access to samples, provided archaeological and osteological context and aided in the interpretation of results. The manuscript was written by D.G.B., C.R., M,-H.S.S. and V.E.M., with contributions from all co-authors.

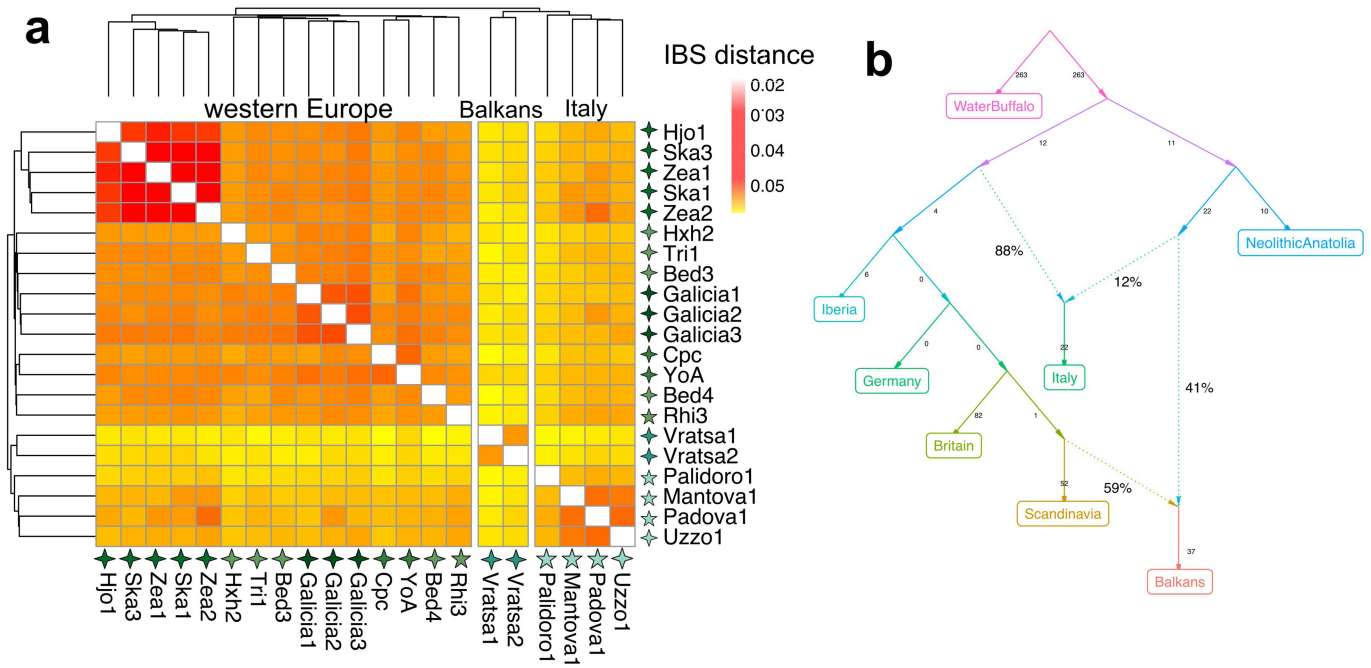
Competing interests The authors declare no competing interests.

Additional information

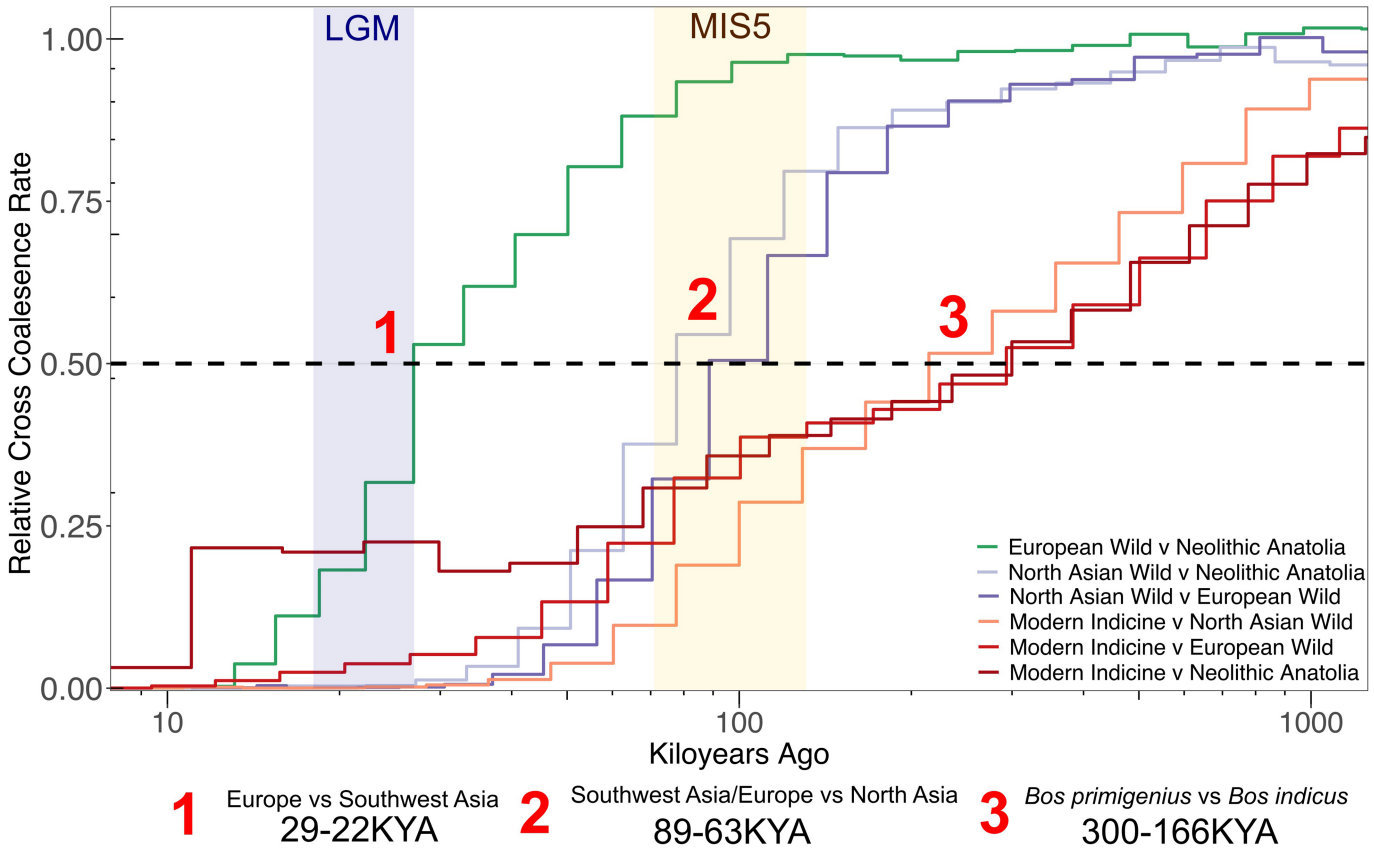
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-08112-6.

Correspondence and requests for materials should be addressed to Mikkel-Holger S. Sinding or Daniel G. Bradley.

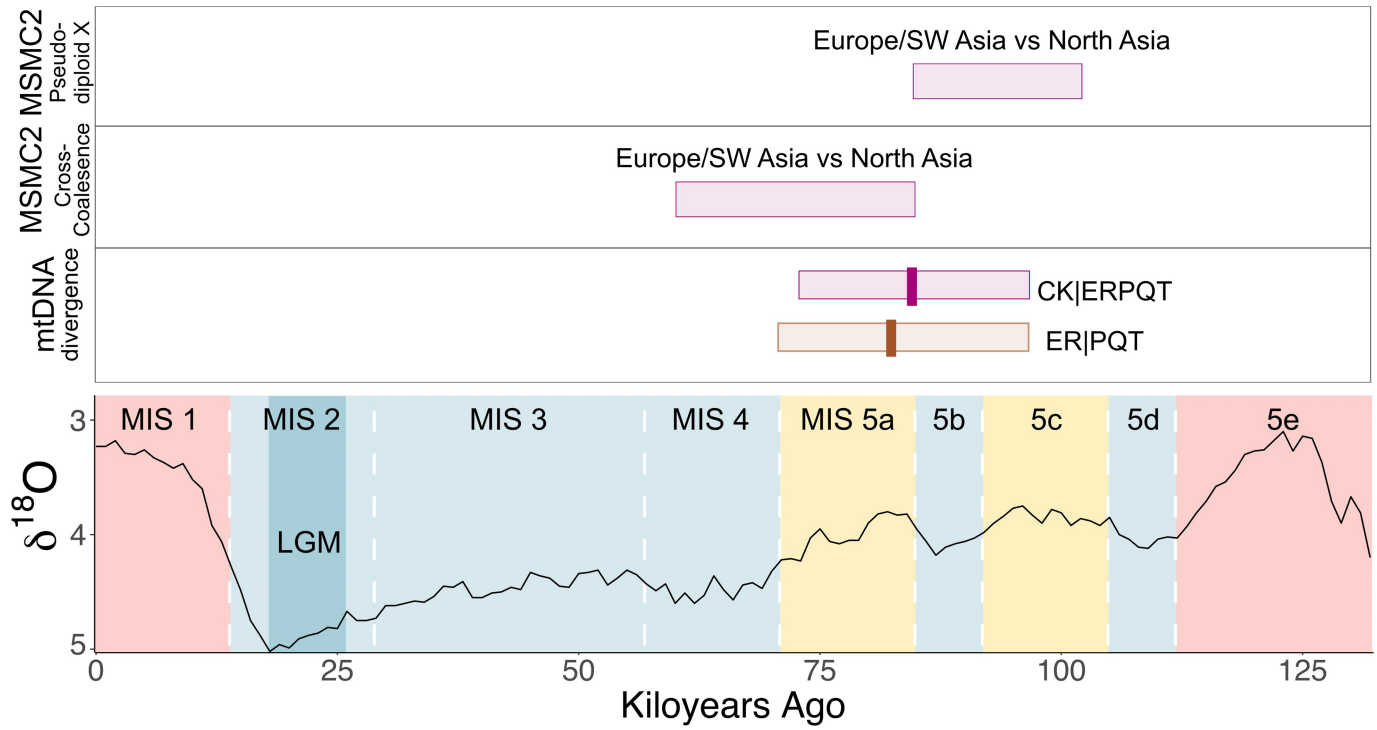
Peer review information *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



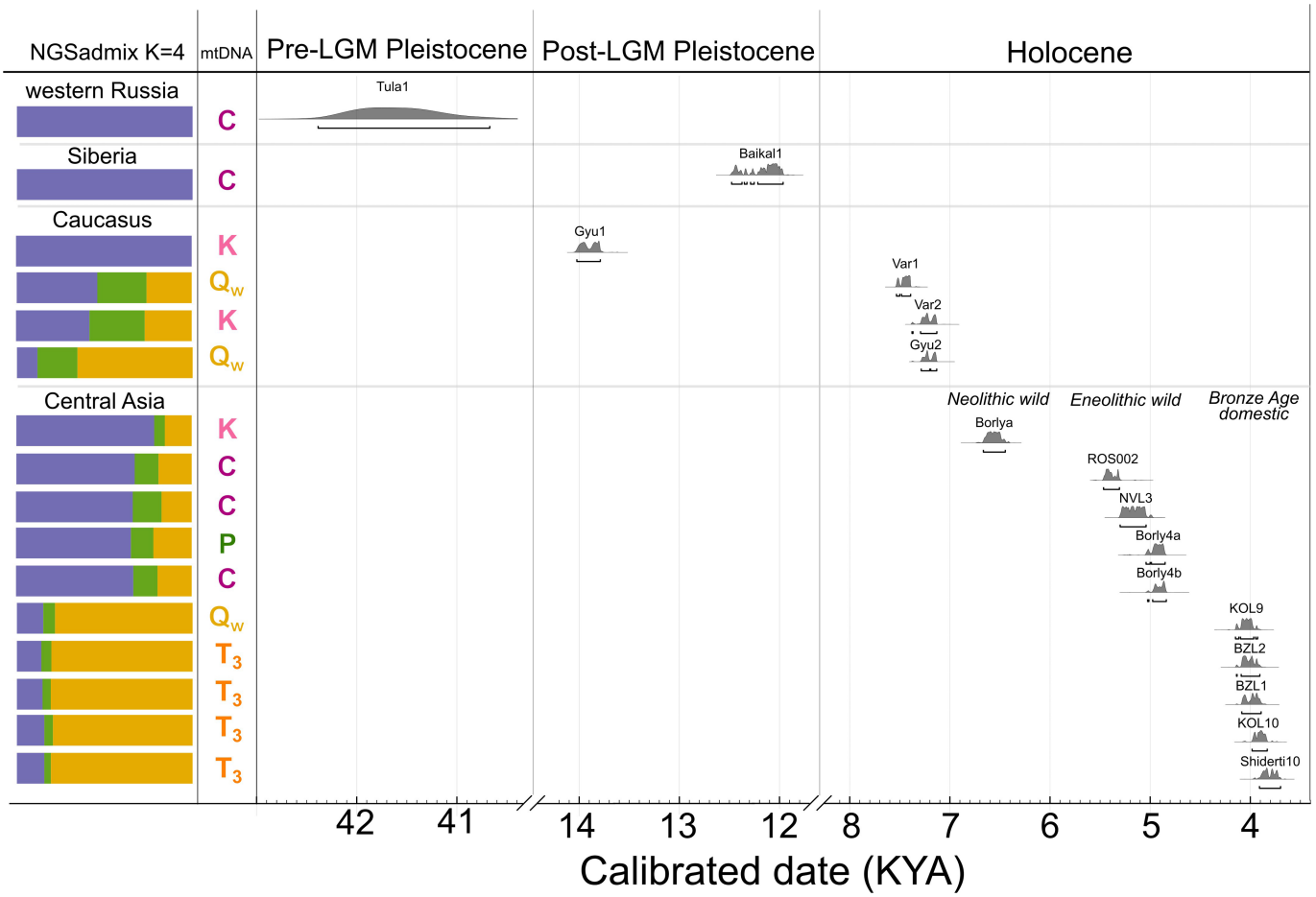
Extended Data Fig. 1 | Three genomic groups within European aurochs. **a**, Heat map of pairwise IBS distance values between European aurochsen drawn using pheatmap (https://github.com/raivokolde/pheatmap). European aurochs cluster into geographically coherent groups; western Europe, Italy and the Balkans. Genomes from Iberia cluster within the western European group, consistent with the Iberian glacial refugium being a source of postglacial recolonization. The lowest pairwise IBS distances are observed in a tight Scandinavian cluster of five genomes in the top left corner. **b**, Best-fitting model estimated using AdmixtureGraph allowing $m = 2$ admixture edges. This infers a closely shared evolutionary history among western European genomes (Iberia, Britain, Germany, Scandinavia). Italians derive from a more basal European source but also have ~12% admixture contribution of a Southwest Asian population. Balkans genomes are modelled as having ~41% Southwest Asian contribution.



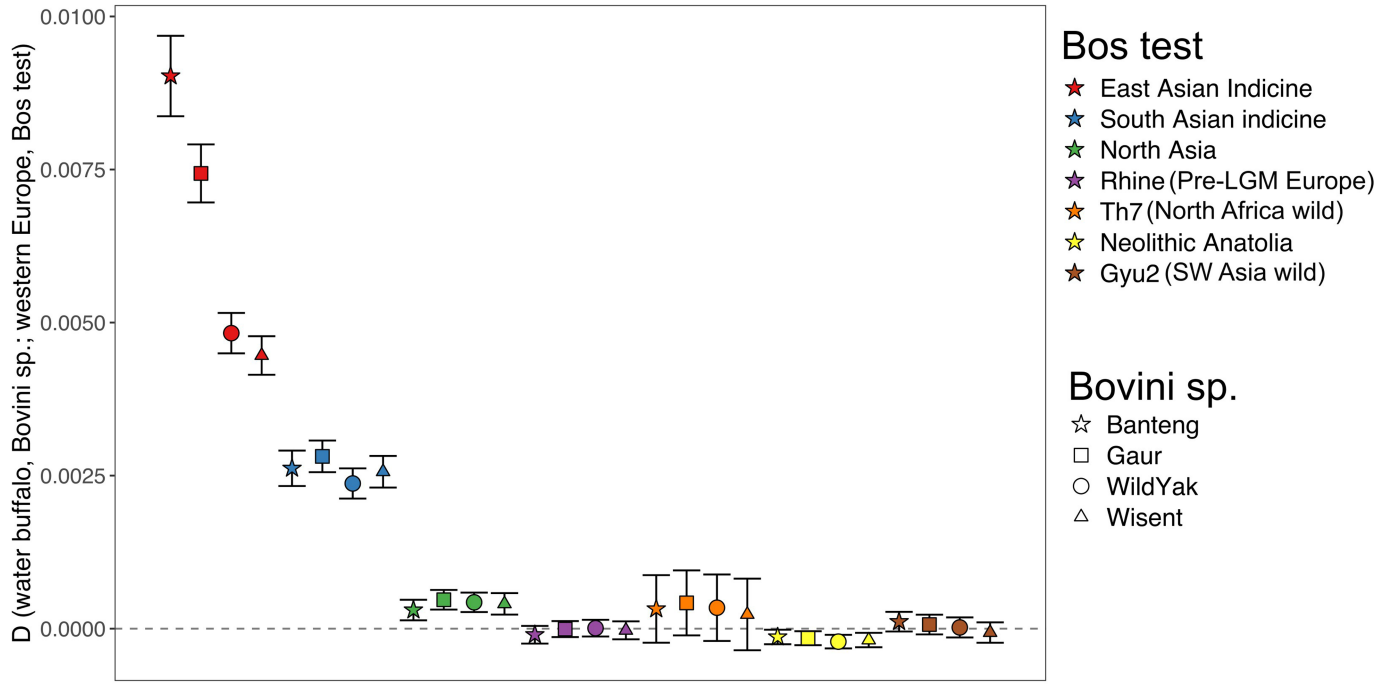
Extended Data Fig. 2 | MSMC2 cross-population analysis. MSMC2 was used to perform cross-population analysis to determine the approximate divergence time between four putative postglacial Eurasian populations; European aurochsen, North Asian aurochsen, Southwest Asian aurochsen (represented by the Anatolian Neolithic, Sub1), South Asian *Bos namadicus* (represented by their descendants, South Asian indicine). The relative cross-coalescence rate (CCR) was computed between all population pairings of the four populations of aurochs and the split time between two populations was estimated as the time interval at the relative CCR of 0.5. Three splits were observed; first European aurochs versus Near Eastern aurochs during the LGM, second European and Southwest Asian/Near Eastern aurochs versus north Asian aurochs during MIS 5, third all *Bos primigenius* populations versus *Bos indicus* (proxy for *Bos namadicus*). The timings of these splits reflect the phylogenetic relationship of the populations.



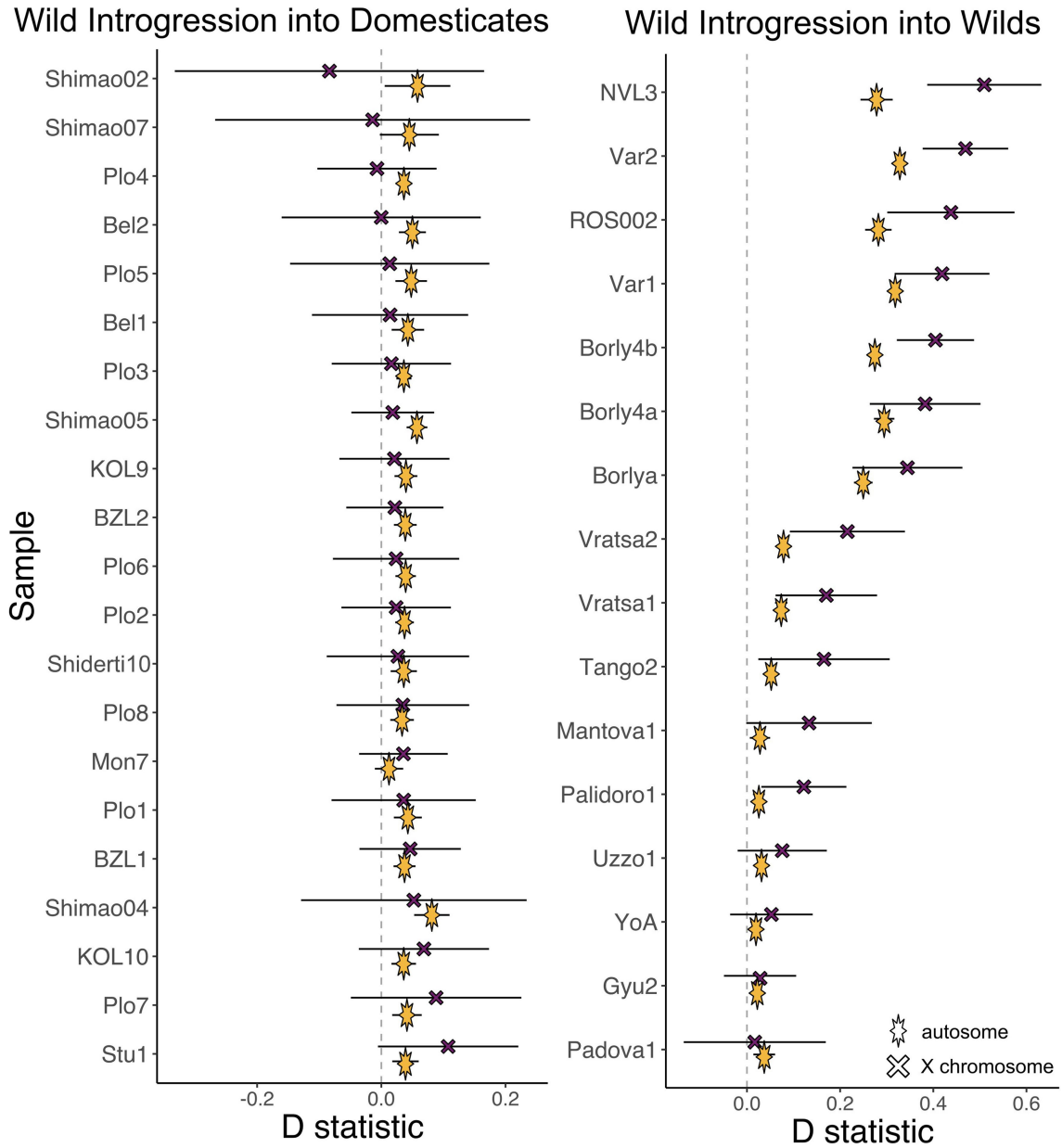
Extended Data Fig. 3 | Summary of analyses timing the divergence of Southwest Asian and European aurochs populations versus North Asian aurochs populations. The top panel displays the range of estimates observed when X chromosome divergence began based on pseudodiploid X chromosome analysis of 10 pairs of North Asian-Southwest Asian/European samples (Supplementary Information Section 4). The second panel displays the range of split time between Southwest Asian/European aurochsen and North Asian aurochsen estimated by MSMC2 cross-population analysis (Extended Data Fig. 2). The third panel displays the 95% HPD interval of major mtDNA haplogroup divergences associated with Southwest Asian/European aurochsen and North Asian aurochsen splits. The final panel displays the LR04 stack of benthic d18O records[89] overlaid with MIS periods[21]. Interglacial periods (MIS 1, MIS 5e) are coloured red, stadial periods (MIS 2-4, MIS 5b, MIS 5d) are coloured blue and interstadial periods (MIS 5a and 5c) are coloured yellow. All divergence estimates overlap with the final interstadial of MIS 5.



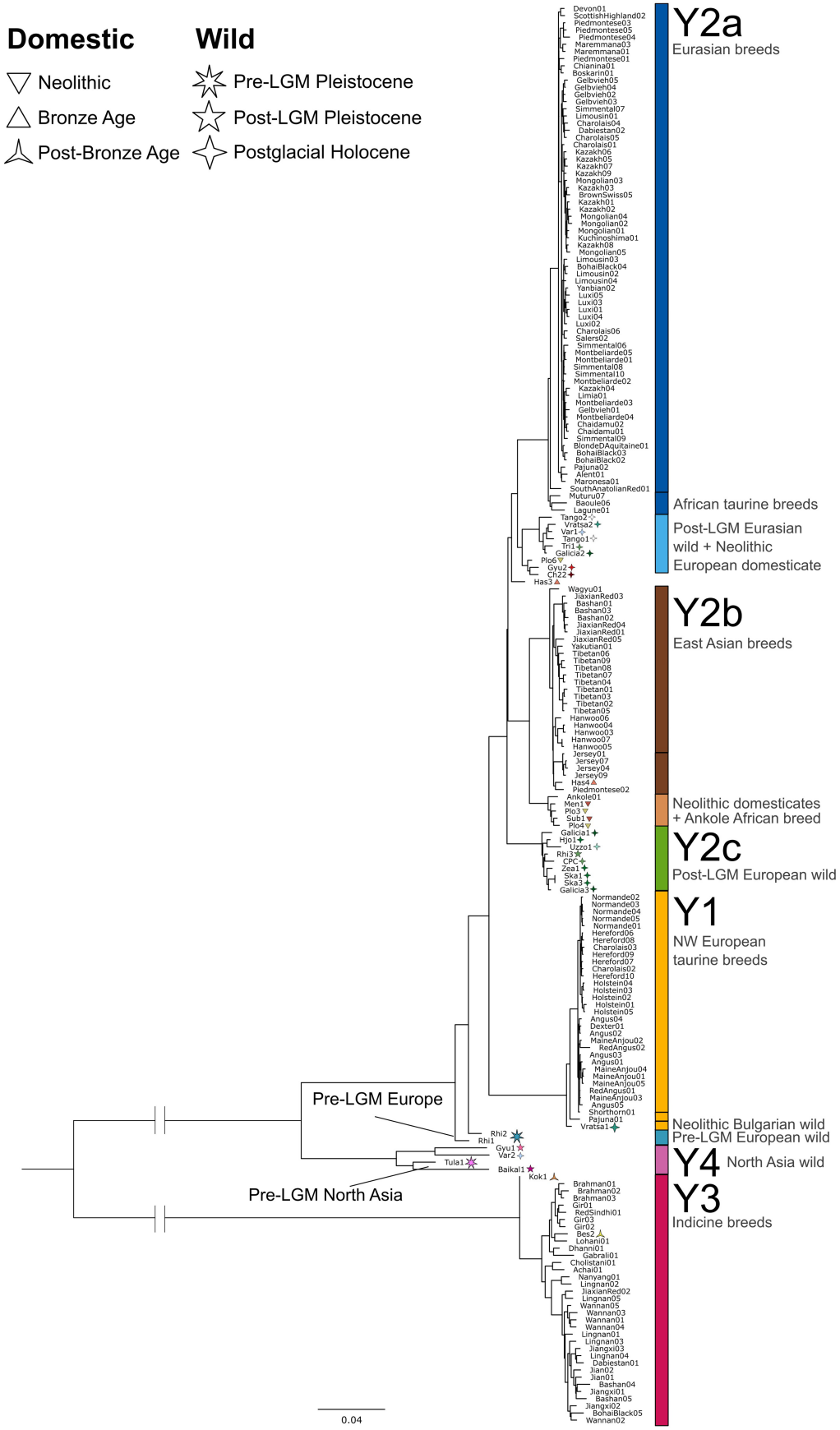
Extended Data Fig. 4 | NGSadmix *K* = 4 composition, mitochondrial haplogroups and ^{14}C dates of 16 Holocene *Bos* sampled from Asia. Pleistocene North Asian samples (Tula1, Baikal1, Gyu1) display an unadmixed ancestral profile that stretches from ~41.5 to ~12.2 ka. By contrast, Holocene Asian aurochsen display an admixed ancestral profile and four separate wild maternal lineages, including the European *P* haplogroup and the Southwest Asian *Q*. The advent of Central Asian domestic taurine cattle during the Bronze Age presents a different ancestral profile, where animals are modelled mostly with the Southwest Asian ancestral component typical of *Bos taurus*, although also with minor North Asian and Europe wild components reflecting introgression from local wild herds (Supplementary Information Section 6). Domestic maternal haplotypes (T3) are detected in 4 of 5 Central Asian domesticates.



Extended Data Fig. 5 | Unequal Bovini allele sharing in *Bos*. We explicitly tested allele sharing of *Bos* populations (East Asian indicine, $n = 3$; South Asian indicine, $n = 4$; North Asia, $n = 3$; Rhine, $n = 2$; North Africa wild, $n = 1$; Neolithic Anatolia, $n = 3$; Southwest Asia wild, $n = 1$). with wild Eurasian Bovini (Banteng, $n = 4$; Gaur, $n = 3$; Wild Yak, $n = 3$; Wisent, $n = 1$) relative to western European aurochs ($n = 12$) using *D* (water buffalo ($n = 3$), Bovini sp.; western Europe, *Bos* test) using SNPs called on Bovini outgroups. We repeatedly observed significant excess allele sharing between indicine and outgroup Bovini species compared to western European populations. All other *Bos primigenius/Bos taurus* populations suggest clade integrity, except Pleistocene North Asia which displayed slightly higher allele sharing with Bovini outgroups, but at an order of magnitude lower than that shown by *Bos indicus*. Error bars depict the standard error of each test multiplied by 3.



Extended Data Fig. 6 | Sex bias in domestic introgression. We explicitly tested allele sharing of *Bos* populations with aurochsen ($n = 15$) relative to Neolithic Anatolian cattle ($n = 3$) with (D (water buffalo ($n = 3$), aurochs population; Neolithic Anatolia, test sample) using two pseudohaploid datasets (SNPs called on the autosome and SNPs called on the X chromosome). In domestic samples ($n = 21$) there is a pattern of increased autosomal contribution relative to the X chromosome, suggesting introgression was male biased. This contrasts with the patterns observed in wild populations, where the minor component is more often female-biased. Error bars depict the standard error of each test multiplied by 3.



Extended Data Fig. 7 | Maximum likelihood phylogeny constructed from Y chromosomal sequences of 206 ancient and modern *Bos* samples. Patterns within the phylogeny reveal that *Bos taurus* and *Bos primigenius* Y chromosomal variation were not reciprocally monophyletic. Haplotypes of ancient and modern domestic cattle were found in six distinct clades and in one solitary individual, suggesting that there were at least seven male contributions to *Bos taurus*.

nature portfolio

Corresponding author(s): Daniel G. Bradley
Mikkel-Holger S. Sinding

Last updated by author(s): **August 28, 2024**

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	cutadapt 3.5, AdapterRemoval v2.3.0, BWA 0.7.5a-r405, samtools v1.13, Picard (version 2.22.1), GATK (v2.4), mapDamage v.2.0.6, github.com/pontussk/PMDtools, Qualimap v2.1.3, Trimmomatic v0.39, bcftools v1.12, ANGSD (version: 0.941-11-g7a5e0db), BEAST v2.6.6, bModelTest 1.2.1, Tracer v1.7.2, vcftools 0.1.17, GLIMPSE v1.1.1, PLINK v1.9, MSMC2 (2.1.3), ape v5.6.2, ADMIXTOOLS2, Clumpak v1.1, Treemix (v1.13), Python (2.7 & 3.5), R (version 4.1.2), PCAngsd (1.01), Integrative Genomics Viewer (1.13.11), NGSadmix (v32), Beagle 5.2, OxCal 4.4,**SeaView (5.05)**

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Sequence reads and alignment files of new data are available through the European Nucleotide Archive under accession number PRJEB75467. Previously published ancient data used in this study are available under accession numbers PRJEB31621, PRJNA379859, PRJNA803479 and PRJEB74338, and detailed in Supplementary Table 2. Sources for previously published modern genomic data analysed here are listed in Supplementary Data Table 3. The ARS-UCD1.2 reference genome is available under NCBI assembly accession GCF_002263795.1. The *Bos taurus* mitochondrial reference genome is available under NCBI accession number V00654.1.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Not applicable
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	DNA was extracted from ancient bovine material. Shotgun sequencing data were generated and aligned to the cattle genome to produce whole genome data and analysed with previously published ancient and modern bovine genomes.
Research sample	50 ancient samples were sequenced in this study, first presented here. These samples range across Eurasia (from Britain to Lake Baikal in Siberia) across a wide temporal range (47+ kya - 3.8 kya). The details of the samples are listed in Supplementary Data 1. These samples were contextualised with other published ancient samples (detailed in Supplementary Data 2) and modern bovinis (detailed in Supplementary Data 3).
Sampling strategy	We sampled 50 new samples from Late Pleistocene to the Holocene, producing 41 whole genomes suitable for phylogenetic analyses (mitochondrial sequences were produced for the remaining 9). The sample size was restricted by the availability of ancient material and the extent of DNA preservation.
Data collection	Ancient genomic data was collected according to laboratory protocols designed to maximise ancient DNA recovery while minimising contamination and destruction of ancient material, detailed in the Methods Section. DNA libraries were sequenced via shotgun sequencing.
Timing and spatial scale	Samples range across Eurasia (from Britain to Lake Baikal in Siberia) across a wide temporal range (47+ kya - 3.8 kya). Sample provenance and radiocarbon age (where available) are reported in Supplementary Data 1.
Data exclusions	All genomes collected for this study were analysed. Thresholds of coverages and subsetting of samples were applied for certain analyses, as detailed in the Methods Section.
Reproducibility	As ancient material finds are rare and unique, it was not possible to reproduce sampling. All bioinformatic analyses were described in detail and all data has been made available, therefore allowing reproduction of analyses by independent groups.

Randomization	Randomisation was not applicable in this study. Samples were grouped into temporal and geographical groups which were relevant to the analysis as detailed in the Methods Section and Supplementary Data 1-3.
Blinding	Blinding of ancient samples was not applicable in this study as all data come from unique archaeological material.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☐	☒ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Palaeontology and Archaeology

Specimen provenance	No new excavations were performed for this study. The age and provenance of the specimens is described in Supplementary Data 1. This table lists the steward institution which was responsible for the storage and curation of each sample. Ancient DNA sampling for each sample is supported with a letter of permission from the steward institution, most of whom are authors on the paper. Where relevant, a Material Transfer Agreement has been made under the The Nagoya Protocol.
Specimen deposition	Supplementary Table 1 lists each sample and its collection identifier. The steward institution that provided access and is responsible for long-term storage and curation of the specimens. Request for access for these specimens may be directed to its respective steward.
Dating methods	New radiocarbon dates were calibrated using the IntCal20 calibration curve using the OxCal v4.4 software. Radiocarbon dating labs are listed in Supplementary Data 1.

☒ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	Sampling was done with the ethical approval of the steward institution.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	not applicable
Novel plant genotypes	not applicable
Authentication	not applicable